POSTER

# Optimal Composition Recommendation for Portrait Photography

**XUE SONG**, Zhengzhou University, Zhengzhou, Henan, China

**JIAWEI PAN**, Beijing University of Posts and Telecommunications, Beijing, Beijing, China

**FUZHANG WU**, Chinese Academy of Sciences, Beijing, Beijing, China

**WEIMING DONG**, Chinese Academy of Sciences, Beijing, Beijing, China

Open Access Support provided by:

Chinese Academy of Sciences

Beijing University of Posts and Telecommunications

Zhengzhou University

**PDF Download**
**3550082.3564206.pdf**
**29 January 2026**
**Total Citations:** 1
**Total Downloads:** 238

**Published:** 06 December 2022

**Citation in BibTeX format**

SA '22: SIGGRAPH Asia 2022
*December 6 - 9, 2022*
*Daegu, Republic of Korea*

**Conference Sponsors:**
SIGGRAPH

SA '22: SIGGRAPH Asia 2022 Posters (December 2022)
https://doi.org/10.1145/3550082.3564206
ISBN: 9781450394628

# Optimal Composition Recommendation for Portrait Photography

Xue Song
Zhengzhou University
Zhengzhou, China
202022592017673@gs.zzu.edu.cn

Jiawei Pan
Beijing University of Posts and Telecommunications
Beijing, China
panjiawei@bupt.edu.cn

Fuzhang Wu[*]
Institute of Software, Chinese Academy Of Sciences
Beijing, China
fuzhang@iscas.ac.cn

Weiming Dong
NLPR, CASIA and Univ.of CAS
Beijing, China
weiming.dong@ia.ac.cn

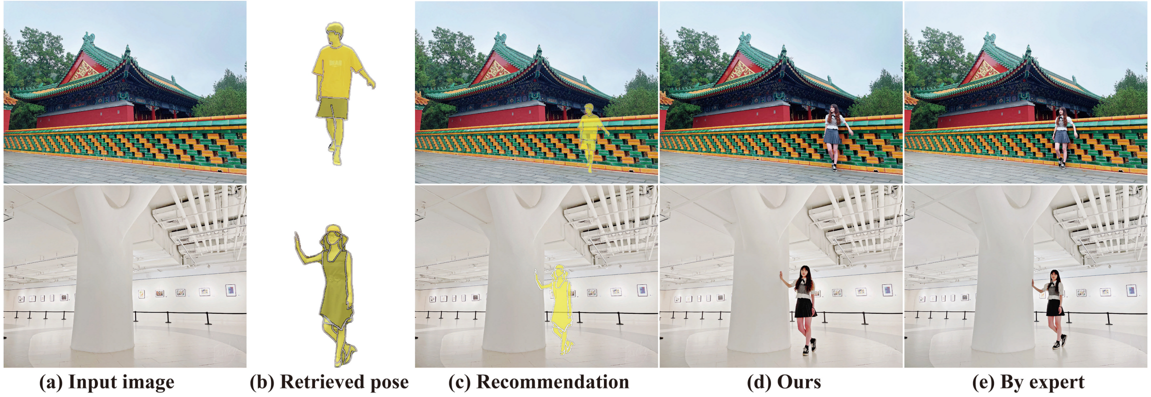

**Figure 1: Given input images (a), we first retrieve suitable poses (b) and predict the composition recommendations (c) by our PACR model. Based on (c), final portrait photos (d) are captured. As a comparison, (e) are taken by an expert.**

**ACM Reference Format:**
Xue Song, Jiawei Pan, Fuzhang Wu, and Weiming Dong. 2022. Optimal Composition Recommendation for Portrait Photography. In . ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3550082.3564206

## 1 INTRODUCTION

Acquiring an appealing photo (especially portrait) needs to consider many aesthetic principles and photo quality assessment attributes. Composition is one of the critical factors that determines the aesthetic quality of a photo. When taking a portrait, professional photographers usually propose a suitable pose and spatial location for the users based on the current scene and general rules of composition. However, this is really challenging for amateur photographers. Wang et al. [2015] proposed a recommendation system to help people take a well-composed portrait picture. However, such system suffers from the same size and position suggestion even if the subjects are with different poses, which is obviously unreasonable.

In this paper, we propose a new composition recommendation algorithm to help users find appropriate poses and search the optimal position and size for the human within the currently framed scene. Our observation is that the composition of a portrait photo is related to the visual content of scene as well as the human posture. Based on this fact, we design an pose-attribute-based composition recommendation (*PACR*) model to learn the composition rules. Inspired by [Su et al. 2021], we formulate the composition recommendation as a position and size adjustment prediction problem. Specifically, given a background image, our system firstly provides a series of appropriate pose candidates based on scene semantic feature retrieval [Tan et al. 2018]. After the user chooses a pose, our algorithm starts from an initial random composition scheme and continues to predict whether the current composition is acceptable or not in aesthetics. If not, our model regresses the adjustment magnitude for each of the candidate adjustments. In our paper, the adjustment operation include vertical or horizontal movement and scaling.

## 2 METHODS

Given a background image $I_B$, we first recommend a proper pose $I_P$ to the user. We build a well-posed portrait dataset (approximately 1000 images in total) and choose a pre-trained CNN-based visual semantic feature extractor [Tan et al. 2018] to retrieval a potential suitable $I_P$ from this dataset. After that, our goal is to predict

[*]Corresponding authors

*Conference Name, Conference Date and Year, Conference Location*
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9462-8/22/12.
https://doi.org/10.1145/3550082.3564206

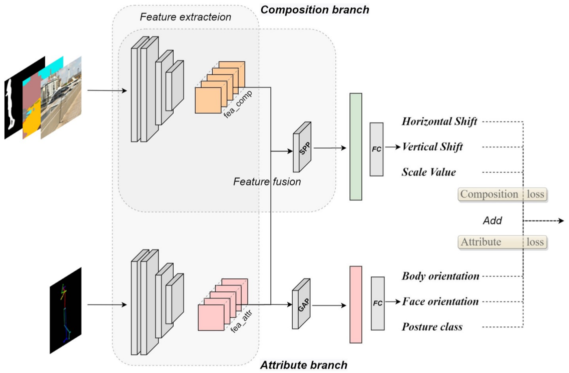

**Figure 2: An overview of the proposed PACR model. The main branch predicts composition adjustment while the auxiliary branch predicts attributes of human pose.**

the well-composed position $(p_x, p_y)$ and size $(p_s)$ of $I_P$. Directly regressing the optimal position and size is extremely challenging. Instead, we propose an optimization strategy to solve this problem. We first define a random initial position and size $P_{init}$=($p_{x_0}$,$p_{y_0}$,$p_{s_0}$) for pose $I_P$, and then predict whether and how to adjust the initial composition. The composition adjustment consists of three terms as defined below:

$$y_{ca} = \begin{cases} \Delta p_x = p_{x_{gt}} - p_{x_0}, \\ \Delta p_y = p_{y_{gt}} - p_{y_0}, \\ \Delta p_s = p_{s_{gt}} - p_{s_0}, \end{cases} \tag{1}$$

where $y_{ca} = (\Delta p_x, \Delta p_y, \Delta p_s)$ is the adjustment magnitude and $p_{(x,y,s)_{gt}}$ indicates the ground truth.

To address the above problem, we develop a dual branch network model (see Fig. 2) which incorporates the composition related features from both background scene and person posture information to predict the adjustment. The main composition adjustment ($CA$) learning branch aims to extract the spatial visual semantic features $\Phi_{CA}(x_{ca})$ of the $I_B$. While the auxiliary pose attributes ($PA$) learning branch is expected to learn the attribute-related composition feature $\Phi_{PA}(x_{pa})$. Finally, we fuse the feature maps $\Phi_{COMP} = (\Phi_{CA}, \Phi_{PA})$ to collectively predict $y_{ca}$.

## 2.1 Composition adjustment (CA) branch

In this branch, we aim to find the underlying relationship between the input images ($I_B$, $I_P$) and the optimal composition. In order to learn a better compostion feature representation, the semantic segmentation $I_{Seg}$ of $I_B$ is also incorporated as the input. We apply a pre-trained CNN-based model to extract spatial semantic feature $\Phi_{CA}(x_{ca})$ from the concatenated input $x_{ca} = (I_B, I_P, I_{Seg})$. Finally, the fused feature maps $\Phi_{COMP}$ are supposed to learn the underlying mapping between $x_{ca}$ and $y_{ca}$. Since this branch predicts a continuous adjustment magnitude $y_{ca}$, we set three nodes in the last layer as the magnitude of adjustment operations, and adopt $L1$ loss function to optimize the task, which can be formulated as:

$$l_{CA} = |y_{ca} - f(x_{ca})|, \tag{2}$$

where $x_{ca}$ is the input image and $y_{ca}$ is the adjustment magnitude label. They are both tensors of shapes $[n,3]$ with a total of $n$ elements each.

## 2.2 Pose attributes (PA) branch

We empirically design three kinds of attributes including body orientation ($attr_b$), face orientation ($attr_f$) and posture ($attr_p$). Each attribute is quaternary: $attr_{b/f} \in$ *{left, right, front, back}* and $attr_p \in$ *{stand, sit, squat, lying / prone}*. We feed the human skeleton image $I_{Skel}$ into a pre-trained CNN-based network and the intermediate feature maps of the model are extracted $\Phi_{PA}(x_{pa} = I_{Skel})$. Cross-entropy loss function is employed to optimize the attribute classification task:

$$l_{attr\_i} = -W_{y_{pa}} log(\frac{e^{x_{pa,y_{pa}}}}{\sum_{c=1}^{C_i} e^{x_{pa,c}}}), \tag{3}$$

where $i$ is the attribute index, $W$ is weight, $x_{pa}$ is the skeleton image, $y_{pa}$ is the attribute labels and $C_i$ is the number of labels for the attribute.

The final loss function of the whole model is defined as:

$$Loss = \sum_{i=1}^{H} l_{attri\_i} + l_{CA}, \tag{4}$$

where $H$ is the number of attributes.

# 3 EXPERIMENTS AND CONCLUSION

Like the work in [Su et al. 2021], we assemble a dataset with approximately 40,000 well-composed portraits from the public aesthetic datasets. All the input images ($I_B$, $I_P$ and $I_{Seg}$) and the pose attribute labels ($y_{pa}$) in our work are generated by the off-the-shelf algorithms (including image matting, inpainting, semantic segmenation and human/face attribute analysis). To improve the label accuracy, additional manual label correction is needed.

Figure 1 shows our composition results. For a fair comparison, a photography expert was invited to take a portrait photo with the same scene and pose idea. We can notice that our composition result ($d$) is visually aesthetic and comparable with expert's work.

In this paper, we propose a dual branch PACR model, which can recommend a well-composed position and size of human instance in a given scene by simultaneously considering the pose attributes and scene semantics. In the future work, we will explore other aesthetic related attributes and extend our work to other types of photography.

# ACKNOWLEDGMENTS

This work was supported by National Science Foundation of China (nos. 61832016, U20B2070).

# REFERENCES

Yu-Chuan Su, Raviteja Vemulapalli, Ben Weiss, Chun-Te Chu, Philip Andrew Mansfield, Lior Shapira, and Colvin Pitts. 2021. Camera view adjustment prediction for improving image composition. *arXiv preprint arXiv:2104.07608* (2021).

Fuwen Tan, Crispin Bernier, Benjamin Cohen, Vicente Ordonez, and Connelly Barnes. 2018. Where and who? automatic semantic-aware person composition. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 1519–1528.

Yinting Wang, Mingli Song, Dacheng Tao, Yong Rui, Jiajun Bu, Ah Chung Tsoi, Shaojie Zhuo, and Ping Tan. 2015. Where2stand: A human position recommendation system for souvenir photography. *ACM Transactions on Intelligent Systems and Technology (TIST)* 7, 1 (2015), 1–22.